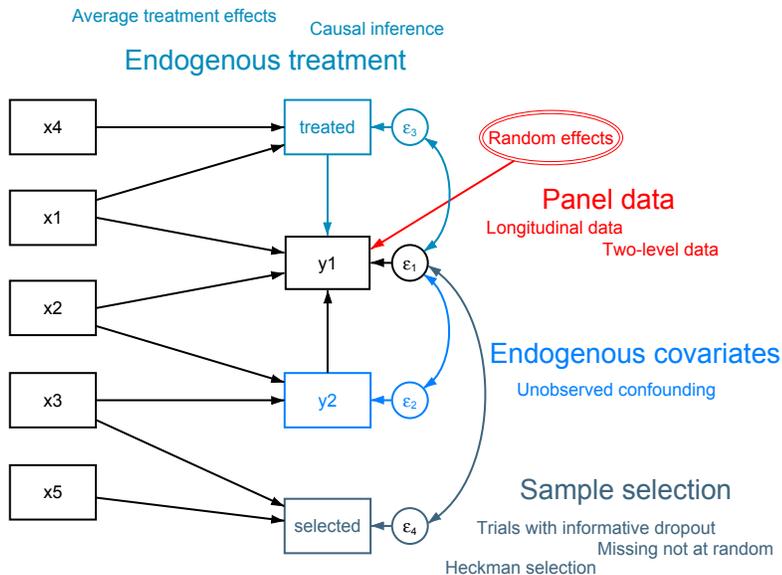


New in

STATA[®] 16

Panel-data extended regression models (ERMs)



ERMs = Endogeneity
+ Selection
+ Treatment
+ Panel data

We call them ERMs—extended regression models. They account for problems that arise in your data, even when those problems arise simultaneously.

Models

- » Outcomes
 - » Continuous
 - » Interval-censored
 - » Binary
 - » Ordinal
- » Maximum likelihood estimation
- » Accounts for
 - » Endogenous covariates
 - » Unobserved confounding
 - » Continuous, binary, and ordinal
 - » Polynomials and interactions
 - » Sample selection
 - » MNAR outcomes
 - » Nonrandom treatment assignment
 - » Casual inference
 - » Endogenous treatment
 - » Exogenous treatment
 - » Panel data
 - » Random effects
 - » Longitudinal data
 - » Two-level data

Advanced inferences

- » Inference statistics
 - » Expected means
 - » Expected probabilities
 - » Contrasts/effects
 - » Marginal effects
 - » Average treatment effects (ATEs)
 - » Potential-outcome means (POMs)
 - » Average structural function (ASF) effects
 - » Average structural probability (ASP) effects
- » Estimates of statistics are available for
 - » Full population
 - » Subpopulations
 - » Means for specified covariate values
 - » Censored and uncensored outcomes
- » Conditional analysis—specify values of all covariates
- » Population-averaged analysis—specify values of some covariates, or none, and average over the rest
- » Tests and confidence intervals for all statistics
- » Profile plots

Panel/longitudinal data

Say you want to fit a model of wages and have repeated data on individuals. You want to use random effects to account for within-panel correlation.

You type

```
. xtregress wage c.age#c.age education
```

Panel data with endogenous covariates

But what if you worry education is endogenous—that unobserved variables that influence education also influence wage? You want to account for the endogenous covariate and within-panel correlation.

You type

```
. xtregress wage c.age#c.age,  
    endogenous(education = med fed)
```

Panel data with sample selection

What if wages are observed only for people who work and you worry that missing outcomes are missing not at random (MNAR)? You want to account for sample selection and within-panel correlation.

You type

```
. xtregress wage c.age#c.age,  
    select(working = age ed minwage)
```

Panel data with nonrandom treatment assignment

What if union membership affects wages and unobserved factors that influence membership also influence wage? You want to account for nonrandom treatment assignment and within-panel correlation.

You type

```
. xtregress wage c.age#c.age,  
    entreat(union = age i.urban i.occupation)
```

Panel data with multiple complications

What if endogenous covariates, sample selection, and nonrandom treatment assignment arise simultaneously in your panel data?

You type

```
. xtregress wage c.age#c.age,  
    endogenous(ed = med fed)  
    select(working = age ed minwage)  
    entreat(union = age i.urban i.occupation)
```

The screenshot shows the Stata command window with the following output:

```
. xtregress wage c.age#c.age, endogenous(education = fed med) nolog
-----
Extended linear regression      Number of obs =      5,899
Group variable: id             Number of groups =    1,000

                                Obs. per group:
                                min =         1
                                avg =        5.1
                                max =         8

Integration method: mvaghermite      Integration pts. =     7
log likelihood = -15981.079          Wald chi(3) =        66.74
                                      Prob > chi2 =         0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
wage					
age	-.010092	.0121382	0.83	0.406	-.0136994 .0338823
c.age#c.age	-.0000733	.0002144	-0.51	0.611	-.0003556 .0002089
education	-.0658868	.0084462	7.80	0.000	-.0824411 -.0493325
_cons	1.895523	.2930888	6.47	0.000	1.322079 2.470966
education					
fed	.2019353	.0025858	120.40	0.000	.237024 .196466
med	-.7991437	.0023521	339.76	0.000	-.7945337 -.8037537
_cons	-.5436336	.0526914	-10.32	0.000	-.646907 -.4403603
var(e.wage)	2.63613	.05828			2.524343 2.752868
var(e.education)	.2053768	.0067358			.2924561 .3188683
corr(e.education,e.wage)	-.1036255	.0156991	6.47	0.000	-.0707682 -.1322885
var(wage[id])	13.32195	.6236351			12.15583 14.59993
var(education[id])	.3917949	.0204081			.3537698 .433907
corr(education[id], wage[id])	-.0823843	.0183698	85.09	0.000	-.0603358 -.091137

Binary, ordinal, or interval-censored outcomes

Panel-data ERM commands work in the same way with other types of outcomes.

If you have a binary outcome, you type

```
. xteprobit highwage c.age#c.age,  
    endogenous(ed = med fed)  
    select(working = age ed minwage)  
    entreat(union = age i.urban i.occupation)
```

If you have an ordinal outcome, you type

```
. xteoprobit wagecat c.age#c.age,  
    endogenous(ed = med fed)  
    select(working = age ed minwage)  
    entreat(union = age i.urban i.occupation)
```

If you have an interval-censored ordinal outcome, you type

```
. xteintreg wagemin wagemax c.age#c.age,  
    endogenous(ed = med fed)  
    select(working = age ed minwage)  
    entreat(union = age i.urban i.occupation)
```

Analyze model results

With any of these models, we can easily analyze the results.

For instance, we can explore wages for different counterfactual levels of age and union membership.

```
. margins union, at(age = (18(5)65))
```

We can easily visualize the results using **marginsplot**.

